

# Tackling Fake News Detection by Continually Improving Social Context Representations using Graph Neural Networks [1]

Nikhil Mehta, Maria Leonor Pacheco, Dan Goldwasser

## Citations

Nikhil Mehta (primary author): 13  
Maria Leonor Pacheco: 115  
Dan Goldwasser: 2542

## Fake News Detection

Fake news is a well-known issue which touches every part of human society [2]. Detecting fake news is a problem which has yet to be solved. There exist publications, such as *The Onion*, which publish satirical news articles. Satire can often be indistinguishable from true fake news. Any model which can detect fake news must also be able to distinguish fake news from satire, which only defines a small part of the problem of detecting fake news. Some publications such as *National Enquirer*, *The Sun*, and others often publish rumors about famous people which is a kind of fake news. Social media companies such as Twitter and Meta likely have an interest in detecting fake news as they have been implicated in the proliferation of fake news.

## Prior Work on Fake News Detection

Detecting fake news has been a topic of numerous papers. Baly [3][4] in several papers proposed a method for labeling entire publications as fake news or not. Some authors have examined fake news by focusing on the users of social media which may have a propensity to spread fake news, other authors focus on articles being labeled as fake news [5], and finally some authors focus on specific claims as being factual or not. In *Predicting Factuality of Reporting and Bias of News Media Sources*, [4] Baly provides an overview of these approaches in the Related Work section.

## Contributions to Fake News Detection

Mehta proposes that fake news may be defined using a graph representation. The proposed graph consists of several layers of nodes: influencers, users, events or articles, and sources. Two types of edges are proposed: observed and inferred. Observed edges come from hard data, such as a user interacting with an event or article. Inferred edges are created by “a set of *inference operators*” proposed by Mehta. Observed edges are added to the graph first, then inferred edges are added.

## Evaluation

The results of the study are mostly an improvement over previous classification studies. Mehta compared their source classification with the results of Baly et al. (2020) [3] and found that the graph built using Mehta's inference operators did better than Baly et al. (2020), though they were not able to directly compare results since Baly has not made their data or methods public. Mehta compared their article classification with Ngyuen et al. [5] and found that their graph representation using inference operators did as much as 4.61% better on average. It should be noted that Mehta found that when attempting to recreate the results of Baly et al. (2020), they were unable to recreate the performance reported by Baly et al. (2020). It would seem that more transparency is required for further studies so that results are more easily reproduced. For their part, Mehta has provided their scraping methodology and their data to the public. Mehta also performed a clustering analysis of new news content and found that their model performed well at detecting fake news and at guessing which users will interact with new content, which could be an indicator of the factuality of the content.

## Impact/Importance

It is not difficult to see the impact that fake news has had on society. A democratic nation depends on its polity to be informed with truthful information. Identifying fake news is an open problem which needs further study and funding. Mehta's approach has markedly improved fake news detection in the datasets they examined. Further work should continue Mehta's example by publishing the methodologies for gathering the data, and their attempts to recreate the results of papers which were less transparent about methodologies. No researcher studying fake news should find themselves in a position to be accused of being fake news themselves. By providing the extensive appendix explaining methodologies that they did, Mehta has set a precedent for how future researchers should publish their results.

## Issues Studying Fake News

Many research papers, including Mehta et al., and Baly et al. (2020), made use of a dataset provided by the website Media Bias/Fact Check (MBFC) to determine if a news source is factual or not. The use of this dataset is problematic, since the gathering of the data is non-scientific. MBFC does not make their data readily available to the public, and as a result researchers perform a web scrape to get the data. The methodology for creating the "Factual Reporting" label is published on the MBFC website thusly:

For each source, a minimum of 10 headlines are reviewed and a minimum of 5 news stories reviewed. We first review news reporting and follow that up by looking at editorial and opinion pieces. If there is still no clear evidence, we will use searching methods for articles on the site such as "Republican," "Democrat," "Liberal," "Conservative," "Trump," "Clinton," etc., until we are sure of political affiliation. This process can be time-consuming or very simple, depending on the source. [6]

Thomas Bennett - trb090020  
CS 4395.001 - UTD Spring '23  
ACL Paper Summary

This methodology seems to be insufficient for scientific purposes. It would seem that further study into how sources are rated as factual or not is needed, though it should be noted that MBFC is doing the best they can as an independent organization.

- [1] Nikhil Mehta, Maria Leonor Pacheco, and Dan Goldwasser. 2022. [Tackling Fake News Detection by Continually Improving Social Context Representations using Graph Neural Networks](#). In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1363–1380, Dublin, Ireland. Association for Computational Linguistics.
- [2] Soroush Vosoughi, Deb Roy, and Sinan Aral. 9 Mar 2018. [The spread of true and false news online](#). In *Science (Volume 359, Issue 6380)*, pages 1146-1151
- [3] Ramy Baly, Georgi Karadzhov, Jisun An, Haewoon Kwak, Yoan Dinkov, Ahmed Ali, James Glass, and Preslav Nakov. 2020. [What Was Written vs. Who Read It: News Media Profiling Using Text Analysis and Social Media Context](#). In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 3364–3374, Online. Association for Computational Linguistics.
- [4] Ramy Baly, Georgi Karadzhov, Dimitar Alexandrov, James Glass, and Preslav Nakov. 2018. [Predicting Factuality of Reporting and Bias of News Media Sources](#). In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 3528–3539, Brussels, Belgium. Association for Computational Linguistics.
- [5] Van-Hoang Nguyen, Kazunari Sugiyama, Preslav Nakov, and Min-Yen Kan. 2022. [FANG: leveraging social context for fake news detection using graph representation](#). *Commun. ACM* 65, 4 (April 2022), 124–132. <https://doi.org/10.1145/3517214>
- [6] Methodology - Media Bias/Fact Check. <https://mediabiasfactcheck.com/methodology/>. Retrieved on Apr 8 2023.